MEASURES OF THE VALUE OF INFORMATION

By John McCarthy

DARTMOUTH COLLEGE, HANOVER, NEW HAMPSHIRE

Communicated by Claude Shannon, July 12, 1956

1. *Introduction.*—Our knowledge of a future event may take the form of a set of probabilities $p_1, \ldots, p_n$. For example, we might have probabilities of $^3/_8$, $^1/_8$, and $^1/_2$ for rain, snow, and clear as tomorrow's weather. In communication theory our interest is in the various events only as carriers of a coded message. For this purpose Shannon's[1] entropy $-\sum p_i \log p_i$ is the appropriate measure of our uncertainty, and a function $A \sum p_i \log p_i + B$ is a good measure of what it is worth to be given these probabilities. In our weather example we care which event occurs. Furthermore, we may be more interested in whether the sky is clear than in whether rain or snow occurs if the weather is bad. In this paper we show that any convex function of a set of probabilities may serve as a measure of the value of information and that two such functions are equivalent in an appropriate sense if and only if they differ by a linear function.

2. *The Forecaster and His Client.*—We get our quantitative measures of the value of information from a situation in which a client pays a forecaster for predictions of a future event according to the following rules:

(i) The forecaster gives the client probabilities $q_1, \ldots, q_n$ for the events, where $\sum q_i = 1$.

(ii) The client takes action on the basis of these probabilities, and one of the possible events occurs.

(iii) If the ith event occurs, the client pays the forecaster $f_i(q_1, \ldots, q_n)$, which is abbreviated $f_i(q)$.

(iv) We assume that neither the forecaster nor the client can influence the predicted event, although the forecaster can make experiments to help predict it, and the client gets an amount which depends on both the action he takes and on the event which occurs. In what follows, it is assumed that the forecaster and the client both wish to maximize the expected value of their incomes.

Assuming that to the forecaster the probabilities of the possible events are $p_1, \ldots, p_n$, his expectation is $\sum p_i f_i(q)$ if he tells the client the q's. A payoff rule is said to "keep the forecaster honest" if, regardless of the value of $p = (p_1, \ldots, p_n)$, the forecaster's expectation is maximized if and only if he puts $q = p$, i.e., $q_i = p_i$ for each i.

Theorem 1. *A payoff rule keeps the forecaster honest if and only if* $f_i(q) = (\partial/\partial q_i)f(q)$, *where* $f(q)$ *is a convex function of* q *which is homogeneous of the first degree. The expectation of an honest forecaster is then* $\sum p_i f_i(p) = f(p)$.

We omit the proof. The derivative has to be taken in a suitable generalized sense. $f(q)$ is called a "payoff function" if it satisfies the conditions of Theorem 1.

I. J. Good[2] considered the problem of paying the forecaster with the restriction that $f_i(q) = F(q_i)$, i.e., the payoff depends only on the probability assigned to the event which actually occurred. He showed that putting $F(x) = A \log x + B$ keeps the forecaster honest, and Gleason (unpublished) showed that this is the only $F(x)$ which does. The forecaster's expectation is then $A \sum p_i \log p_i + B$, i.e., he is paid a fixed fee *minus* the expected uncertainty about the event after his prediction.

3. *The Client's Expectation.*—Suppose that on the basis of the forecaster's prediction the client chooses the jth of the actions open to him and that his payoff if the ith event occurs is a_{ij}. His expectation will be $g(p) = \max \sum_i a_{ij}p_i$ if j is chosen optimally.

From the theory of convex functions we have

THEOREM 2. *Any function $g(p)$ defined for $p_1 \geq 0, \ldots, p_n \geq 0$ which is convex and homogeneous of the first degree can be written in the form* $\max_j \sum a_{ij}p_i$. *Unless $g(p)$ is piecewise linear, there will have to be an infinite number of actions j.*

If we put $f(p) = g(p)$, the client is eliminated from the picture, since under this condition he turns all his gains over to the forecaster and is reimbursed for all his losses. This is not a satisfactory solution to the problem, so let us see what payoffs f are equivalent in their effect on the forecaster's efforts to get information.

4. *The Forecaster's Experiments.*—Assume that the forecaster has a priori probabilities $r_1, \ldots, r_n$ for the events, that he has a choice of m experimental procedures with expected costs to him of $c_1, \ldots, c_m$, and that the conditional probability of the kth outcome of the hth experiment given that the ith event will occur is s_{khi}. The experiment chosen by the forecaster will depend on the c's, the s's, and the r's and on the payoff function chosen by the client. We call two payoff rules equivalent if, for any set of c's, s's, and r's, they lead to the same choice of experiment by the forecaster.

THEOREM 3. *$f(q)$ and $f^*(q)$ are equivalent if and only if $f(q) = f^*(q) + \sum a_i q_i$, i.e., if the two payoff functions differ by a linear function of the q's.*

The proof is omitted. If f and f^* are equivalent, then $f_i(q) = f_i^*(q) + a_i$, so that the payoff rules differ by an amount which depends only on the event which occurs and not on the forecaster's prediction. The forecaster's and client's interests will be identical if we put $f(q) = g(q) + \sum a_i q_i$. The a_i's are subject to negotiation between the client and the forecaster, and they determine both a base level of payment and also a betting relation between the client and forecaster. If f is normalized so that $f(1, 0, \ldots, 0) = f(0, 1, \ldots, 0) = \ldots$, the payment for a precise correct prediction is independent of the event predicted.

5. *Conclusion.*—The foregoing analysis shows that any convex function of a set of probabilities will, under appropriate circumstances, be a measure of the value of the information contained in a set of probabilities in the sense that it is an appropriate payment to a forecaster who furnishes the probabilities.

The intuitive content of the convexity restriction is that it is always a good idea to look at the outcome of an experiment if it is free. For suppose that the experiment has two outcomes, A and A^*, which would give one probabilities p and p^* for the event in question. Let t be the probability that A is the outcome. If we decide not to look, our expectation is $f(tp + (1 - t)p^*)$, while if we decide to look, our expectation is $tf(p) + (1 - t)f(p^*)$.

Finally, we remark that there are yet more general ways of paying the forecaster. For example, the client may agree to pay a certain fraction α of the costs of experimentation. Then the payoff function can be scaled down by a factor α with the identity of interests still preserved. We hope to treat these matters on another occasion.

[1] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (Urbana: University of Illinois Press, 1949).

[2] I. J. Good, "Rational Decisions," *J. Roy. Stat. Soc., B*, Vol. **14**, No. 1, 1952.